


[Subscribe \(Full Service\)](#) [Register \(Limited Service, Free\)](#) [Login](#)
**Search:**  The ACM Digital Library  The Guide



## THE ACM DIGITAL LIBRARY

[Feedback](#)

Web page categorization      Found 1,952  
 Terms used:  
**Web page categorization**      of 241,122

Sort results by  
 Display results

relevance  
 expanded form

Save these results to a Binder  
[Advanced](#)

Open results in a new window  
[Search](#)  
 Try this search in [The ACM Guide](#)

---

Results 1 - 20 of 1,952 Result page: 1 [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [next](#)

[>>](#)

### [1 Automatic web pages categorization with ReliefF and Hidden Naive Bayes](#)



Xin Jin, Rongyan Li, Xian Shen, Rongfang Bie

March 2007 SAC '07: Proceedings of the 2007 ACM symposium on Applied computing

2007

**Publisher:** ACM

Full text available: [pdf\(117.40 KB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

A great challenge of web mining arises from the increasingly large web pages and the high dimensionality associated with natural language. Since classifying web pages of an interesting class is often the first step of mining the web, web page categorization/classification ...

**Keywords:** Hidden Naive Bayes, ReliefF feature selection, web mining

## 2 Linear prediction models with graph regularization for web-page categorization

 Tong Zhang, Alexandrin Popescul, Byron Dom

August 2006 KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining

**Publisher:** ACM

Full text available:  pdf(747.54 KB)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

We present a risk minimization formulation for learning from both text and graph structures which is motivated by the problem of collective inference for hypertext document categorization. The method is based on graph regularization formulated as a well-formed ...

**Keywords:** collective inference, document classification, graph and relational learning, regularization, semi-supervised learning

## 3 Machine learning methods for Chinese web page categorization

Ji He, Ah-Hwee Tan, Chew-Lim Tan

October 2000 Proceedings of the second workshop on Chinese language

processing: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics - Volume 12, Volume 12

**Publisher:** Association for Computational Linguistics

Full text available:  pdf(706.21 KB)

Additional Information: [full citation](#), [abstract](#), [references](#), [cited by](#)

This paper reports our evaluation of  $k$  Nearest Neighbor (kNN), Support Vector Machines (SVM), and Adaptive Resonance Associative Map (ARAM) on Chinese web page classification. Benchmark experiments based on a Chinese web corpus showed that their ...

## 4 Web-page classification through summarization

 Dou Shen, Zheng Chen, Qiang Yang, Hua-Jun Zeng, Benyu Zhang, Yuchang Lu, Wei-Ying Ma

July 2004 SIGIR '04: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval

**Publisher:** ACM

Full text available:  pdf(225.80 KB)

Additional Information: [full citation](#), [abstract](#), [references](#), [cited by](#), [index terms](#), [review](#)

Web-page classification is much more difficult than pure-text classification due to a large variety of noisy information embedded in Web pages. In this paper, we propose a new Web-page classification algorithm based on Web summarization for improving ...

**Keywords:** content body, web page categorization, web page summarization

## 5 Estimating the evolution of categorized web page populations

 Ioannis Anagnostopoulos, Photis Stavropoulos, Georgios Kouzas, Christos Anagnostopoulos, Dimitrios D. Vergados

July 2006 ICWE '06: Workshop proceedings of the sixth international conference on Web engineering

**Publisher:** ACM

Full text available:  pdf(112.13 KB)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

This paper proposes a statistical approach for estimating the evolution of categorized web pages. The proposal is based on the capture-recapture method used in wildlife biological studies and it is modified according to the necessary assumptions and ...

**Keywords:** capture-recapture measurements, web evolution, web page categorization

## 6 Knowing a web page by the company it keeps

 Xiaoguang Qi, Brian D. Davison

November 2006 CIKM '06: Proceedings of the 15th ACM international conference on Information and knowledge management

**Publisher:** ACM

Full text available:  pdf(501.49 KB)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Web page classification is important to many tasks in information retrieval and web mining. However, applying traditional textual classifiers on web data often produces unsatisfying results. Fortunately, hyperlink information provides important clues ...

**Keywords:** SVM, neighboring, rainbow, web page classification

## 7 Mining web content outliers using structure oriented weighting techniques and N-grams



Malik Agyemang, Ken Barker, Rada S. Alhajj

March 2005 SAC '05: Proceedings of the 2005 ACM symposium on Applied computing

**Publisher:** ACM

Full text available:  pdf(106.61 KB)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Classifying text into predefined categories is a fundamental task in information retrieval (IR). IR and web mining techniques have been applied to categorize web pages to enable users to manage and use the huge amount of information available on the ...

**Keywords:** dissimilarity measure, n-grams, text categorization, web contents, web mining

## 8 Topic Continuity for Web Document Categorization and Ranking

B. L. Narayan, C. A. Murthy, Sankar K. Pal

October W1 '03: Proceedings of the 2003 IEEE/WIC International Conference on 2003 Web Intelligence

**Publisher:** IEEE Computer Society

Full text available:  [Publisher Site](#)

Additional Information: [full citation](#), [abstract](#), [cited by](#), [index terms](#)

PageRank is primarily based on link structure analysis. Recently, it has been shown that content information can be utilized to improve link analysis. We propose a novel algorithm that harnesses the information contained in the history of a surfer to ...

## 9 Interest-based personalized search

 Zhongming Ma, Gautam Pant, Olivia R. Liu Sheng

February 2007 ACM Transactions on Information Systems (TOIS), Volume 25 Issue 1

**Publisher:** ACM

Full text available:  [pdf\(1.58 MB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Web search engines typically provide search results without considering user interests or context. We propose a personalized search approach that can easily extend a conventional search engine on the client side. Our mapping framework automatically maps ...

**Keywords:** Open Directory, Personalized search, World Wide Web, information retrieval, user interest, user interface

## 10 Joint categorization of queries and clips for web-based video search

 Ruofei Zhang, Ramesh Sarukkai, Jyh-Herng Chow, Wei Dai, Zhongfei Zhang

October 2006 MIR '06: Proceedings of the 8th ACM international workshop on Multimedia information retrieval

**Publisher:** ACM

Full text available:  [pdf\(378.76 KB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Building a video search engine on the Web is a very challenging problem. Compared with web page search, video search has its unique characteristics (such as high volume of data for each video, existence of multi-modal information including meta-data, ...

**Keywords:** experiment, multi-modality based categorization, query categorization, video categorization, web-based video search

## 11 Categorizing web queries according to geographical locality

Luis Gravano, Vasileios Hatzivassiloglou, Richard Lichtenstein  
November 2003 CIKM '03: Proceedings of the twelfth international conference on Information and knowledge management

**Publisher:** ACM

Full text available:  pdf(545.74 KB)

Additional Information: [full citation](#), [abstract](#), [references](#), [cited by](#), [index terms](#)

Web pages (and resources, in general) can be characterized according to their *geographical locality*. For example, a web page with general information about wildflowers could be considered a *global* page, likely to be of interest to a geographically ...

**Keywords:** information retrieval, query classification, query modification, search engines, web search

## 12 The Role of URLs in Objectionable Web Content Categorization

Jianping Zhang, Jason Qin, Qiuming Yan  
December 2006 WI '06: Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence

**Publisher:** IEEE Computer Society

Full text available:  pdf(233.52 KB)

Additional Information: [full citation](#), [abstract](#), [index terms](#)

By analyzing a set of access attempts by teenagers to pornographic websites, we found that more than half of them are image searches and visits to websites with little text information. It is obvious that textual content-based filters cannot correctly ...

## 13 Measuring Semantic Similarity between Named Entities by Searching the Web Directory

Jiahui Liu, Larry Birnbaum  
November 2007 WI '07: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence

**Publisher:** IEEE Computer Society

Full text available:  pdf(525.06 KB)

Additional Information: [full citation](#), [abstract](#)

The importance of named entities in information retrieval and knowledge management has recently brought interest in characterizing semantic relationships between entities. In this paper, we propose a method for measuring semantic similarity, an important ...

**14 An unsupervised hierarchical approach to document categorization**

Robert Wetzker, Tansu Alpcan, Christian Bauckhage, Winfried Umbrath, Sahin Albayrak  
November WI '07: Proceedings of the IEEE/WIC/ACM International Conference on  
2007 Web Intelligence

**Publisher:** IEEE Computer Society

Full text available:  pdf(401.04)  


Additional Information: [full citation](#), [abstract](#)

We propose a hierarchical approach to document categorization that requires no pre-configuration and maps the semantic document space to a predefined taxonomy. The utilization of search engines to train a hierarchical classifier makes our approach more ...

**15 A Report of Activities at the WIC-India Research Center**

B. L. Narayan, Sankar K. Pal  
September WI '04: Proceedings of the 2004 IEEE/WIC/ACM International  
2004 Conference on Web Intelligence

**Publisher:** IEEE Computer Society

Full text available:  pdf(63.57 KB)  Publisher  
 Site

Additional Information: [full citation](#), [abstract](#)

The research activities of the WIC-India Research Center include topics like improving the performance of search engines, link and neighborhood analysis, as well as, surfer modeling for ranking and categorization of web pages, and query answering. In ...

**16 What type of page is this?: genre as web descriptor**

 Mark A. Rosso  
June JCDL '05: Proceedings of the 5th ACM/IEEE-CS joint conference on Digital  
2005 libraries

**Publisher:** ACM

Full text available:  pdf(135.20)  


Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Many have suggested the use of genres to ameliorate the problem of web search, e.g. [1,3,4,5,6,7]. A central issue in the implementation of this idea is the choice of genres to be used as web page descriptors. Several studies have explored user terminology ...

**Keywords:** classification, genre, metadata, web search

**17 Analysis of usage patterns in experiential multiple perspective web search**

◆ **Rahul Singh, Ya-Wen Hsu**  
September 2007 **MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia**  
**Publisher:** ACM  
Full text available:  [pdf\(868.40 KB\)](#) **Additional Information:** [full citation](#), [abstract](#), [references](#), [index terms](#)

With the rapid growth in the volume, complexity, and heterogeneity of information in the World Wide Web (WWW), the role of user-data interaction paradigms is becoming increasingly critical to the success of web-based information retrieval and assimilation. ...

**18 Accuracy enhancement of function-oriented web image classification**

◆ **Koji Nakahira, Toshihiko Yamasaki, Kiyoharu Aizawa**  
May 2005 **WWW '05: Special interest tracks and posters of the 14th international conference on World Wide Web**  
**Publisher:** ACM  
Full text available:  [pdf\(299.81 KB\)](#) **Additional Information:** [full citation](#), [abstract](#), [references](#), [index terms](#)

We propose a function-oriented classification of web images and show new applications using this categorization. We defined nine categories of images taking into account of their functions used in web pages, and classified web images by using Support ...

**Keywords:** classification, support vector machine, web images

**19 Improved annotation of the blogosphere via autotagging and hierarchical clustering**

◆ **Christopher H. Brooks, Nancy Montanez**  
May 2006 **WWW '06: Proceedings of the 15th international conference on World Wide Web**  
**Publisher:** ACM  
Full text available:  [pdf\(281.51 KB\)](#) **Additional Information:** [full citation](#), [abstract](#), [references](#), [index terms](#)

Tags have recently become popular as a means of annotating and organizing Web pages and blog entries. Advocates of tagging argue that the use of tags produces a 'folksonomy', a system in which the meaning of a tag is determined by its use among the community ...

**Keywords:** automated annotation, blogs, hierarchical clustering, tagging

**20** [Improving web performance by client characterization driven server adaptation](#)



Balachander Krishnamurthy, Craig E. Wills

May 2002 WWW '02: Proceedings of the 11th international conference on World Wide Web

**Publisher:** ACM

Full text available: [pdf\(241.76 KB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [cited by](#), [index terms](#)

We categorize the set of clients communicating with a server on the Web based on information that can be determined by the server. The Web server uses the information to direct tailored actions. Users with poor connectivity may choose not to stay at ...

**Keywords:** client characterization, client connectivity, server adaptation

---

Results 1 - 20 of 1,952 Result page: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [next](#)

[>>](#)

The ACM

Portal is published by the Association for Computing Machinery. Copyright © 2008 ACM, Inc.

[Terms of Usage](#) [Privacy Policy](#) [Code of Ethics](#) [Contact Us](#)

Useful downloads:

[Adobe Acrobat](#)

[QuickTime](#)

[Windows Media Player](#)

[Real Player](#)